

Virginia Highways: MAP-21

by Christian Schroeder, Timothy (Will) Tyree, Colin Warner



Executive Summary

Research Question: Is it possible to predict the reliability of travel time for each Virginia interstate segment?

Response Variable:

RELIABLE

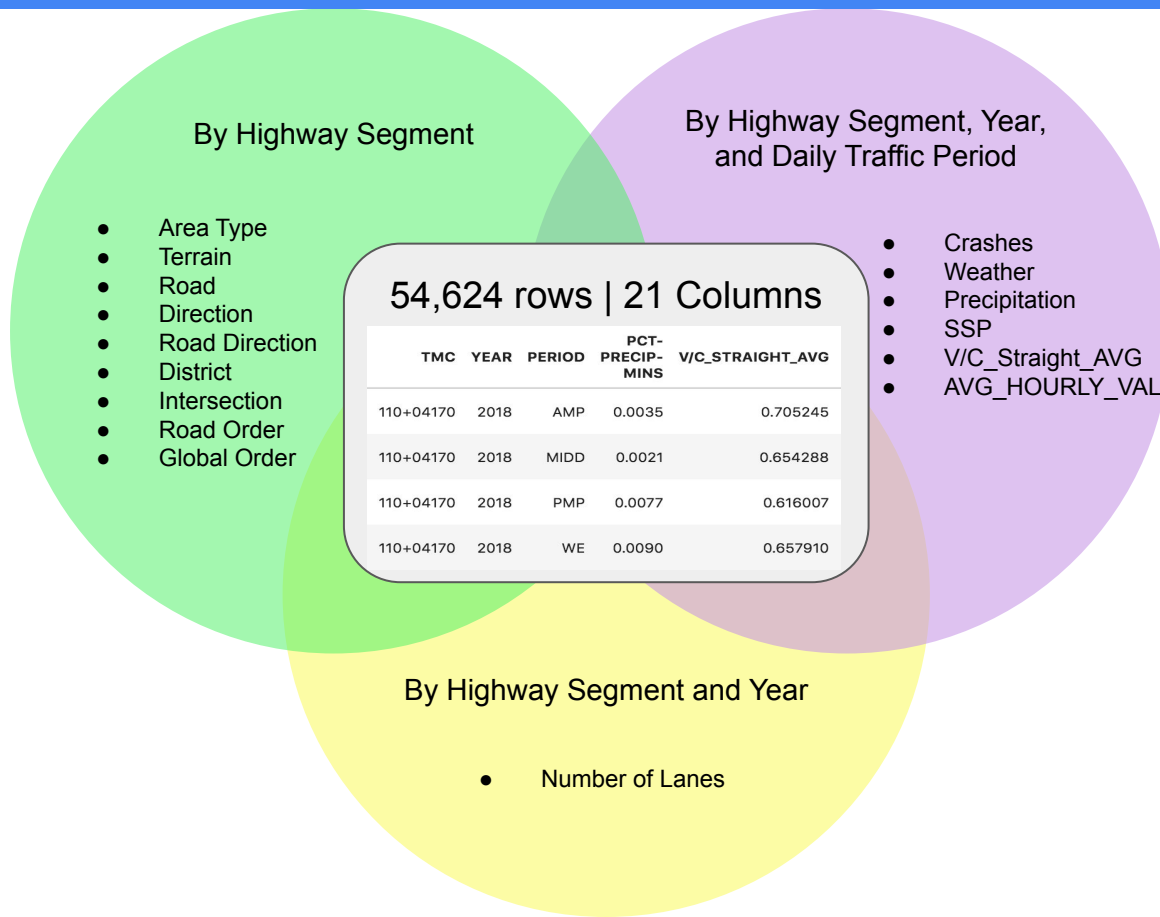
OR

UNRELIABLE

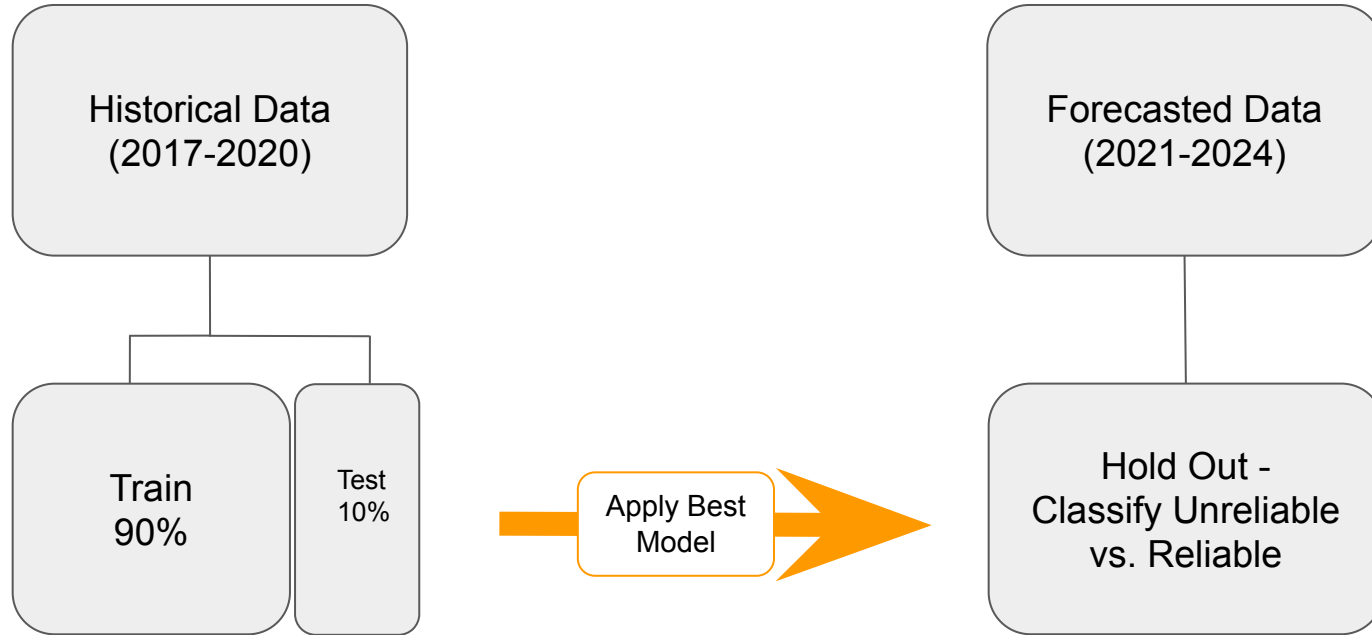
Findings:

Decision Tree accurately identifies unreliable segments for VDOT.

Data Import and Preprocessing



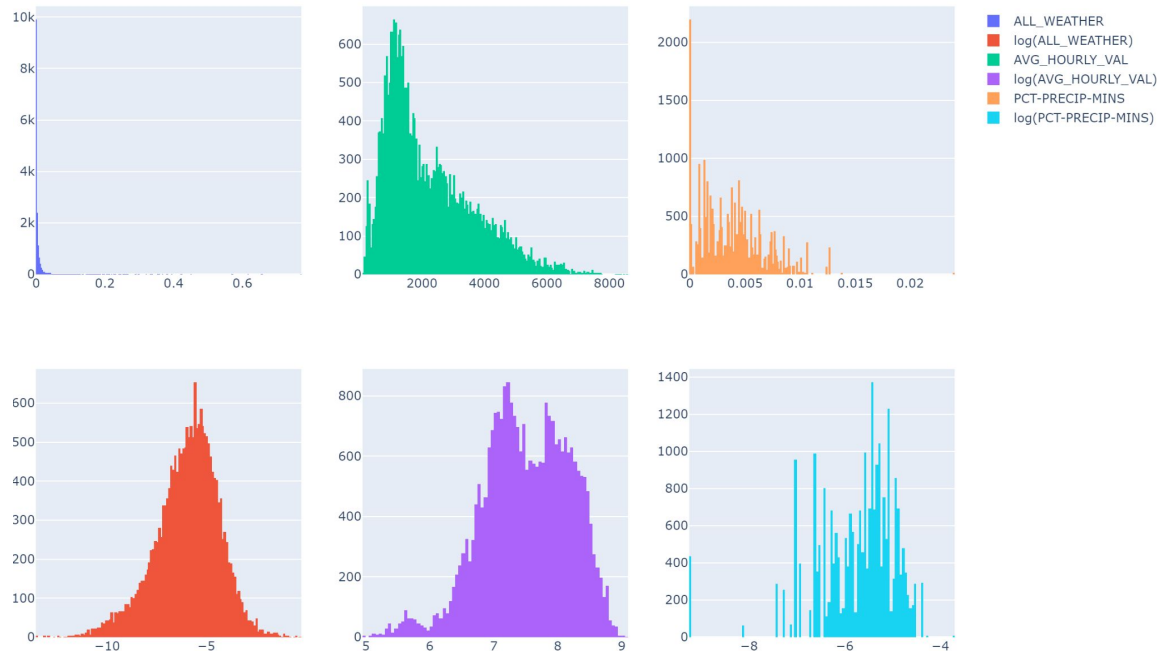
Data Splitting



Exploratory Data Analysis

Check distribution of numeric variables - do they benefit from Log transformation?

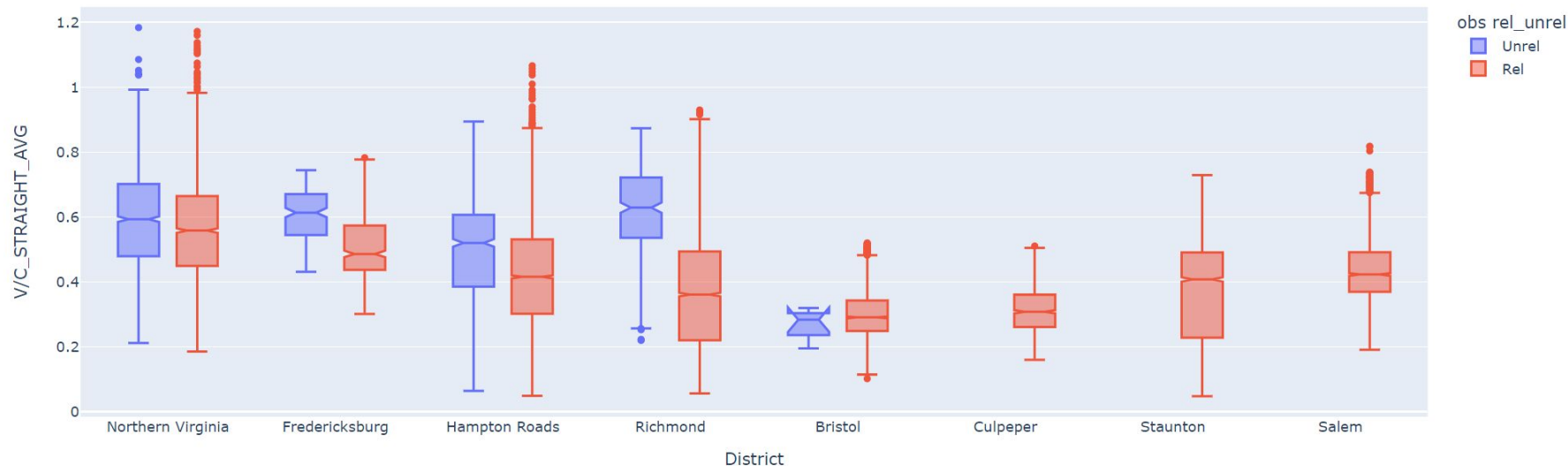
Weather Events, Hourly Volume Rate, and Precipitation Rate Distribution vs. Log Transform Distribution



Exploratory Data Analysis

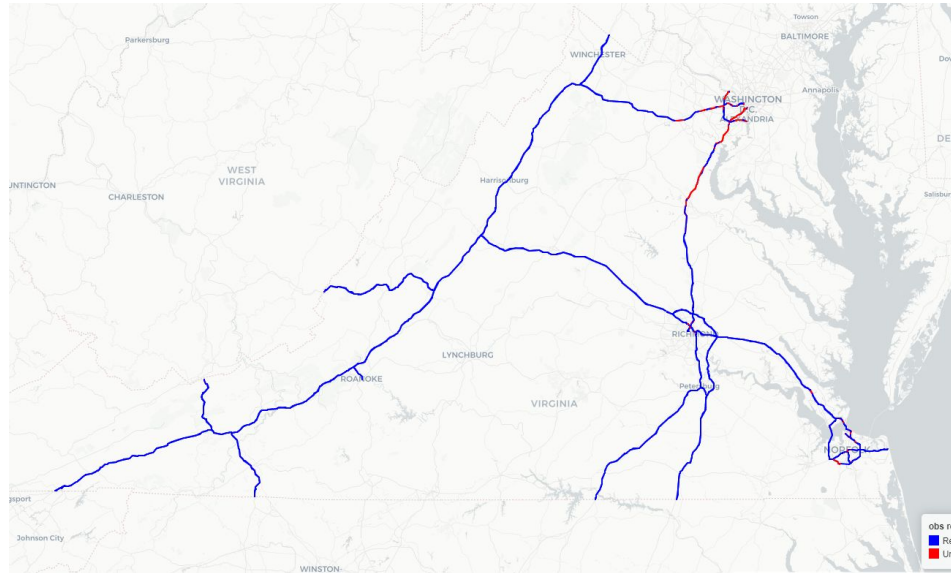
How do Volume/Capacity and District variables impact reliability?

How District and Volume Affect Reliability

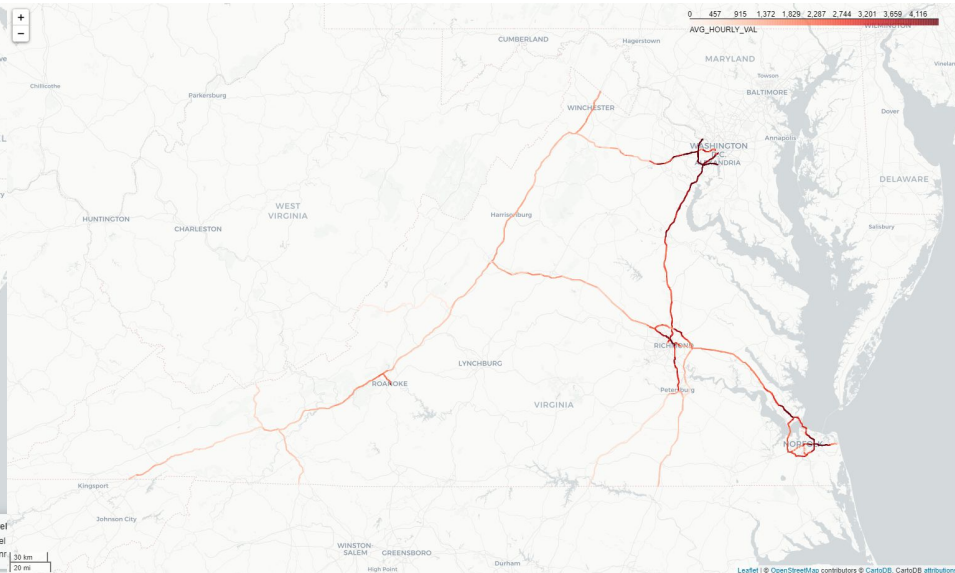


Exploratory Data Analysis

2017 - Reliability

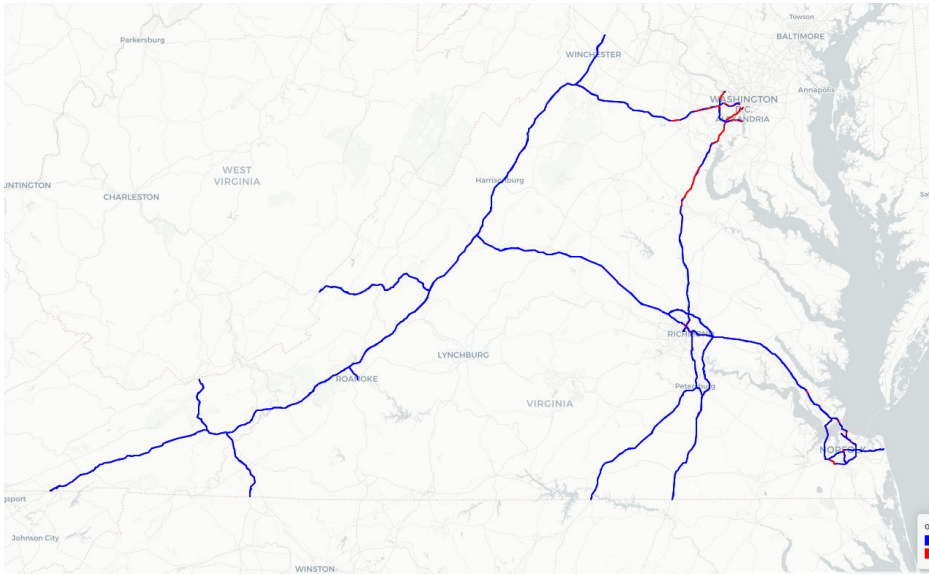


2017 - Hourly Traffic Volume

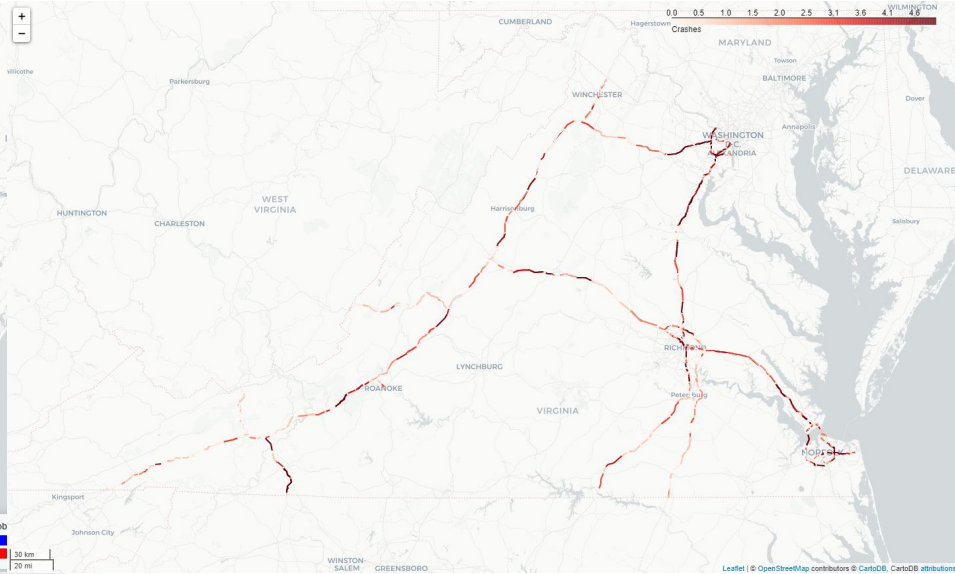


Exploratory Data Analysis

2017 - Reliability

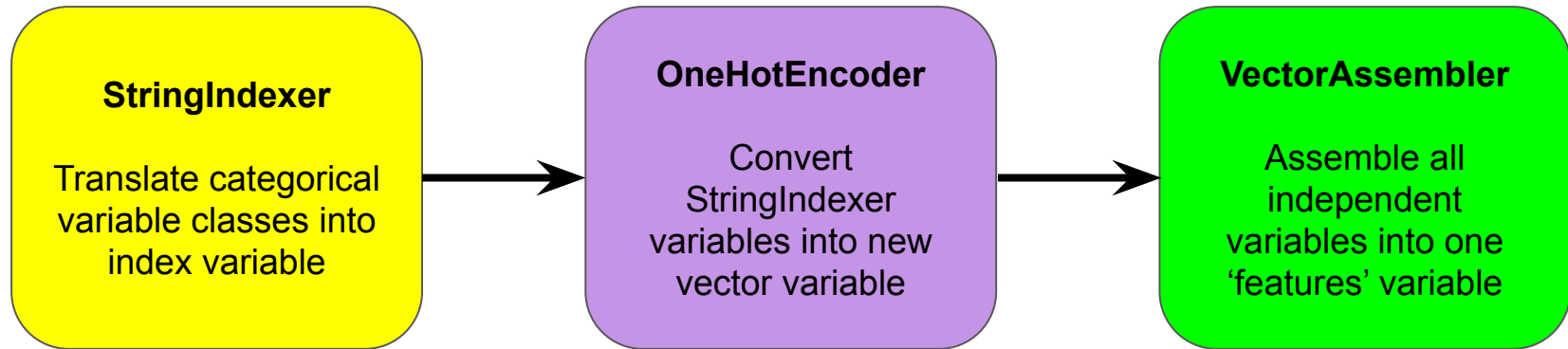


2017 - Crashes



Model Construction | Build Pipeline

```
pipeline = Pipeline(stages=[stringIndexer, encoder, assembler])
```



Model Construction | Logistic Regression

Parameter Grid

```
= (ParamGridBuilder() \  
  .addGrid(lr.regParam, [0.01, 0.5, 2.0]) \  
  .addGrid(lr.elasticNetParam, [0.0, 0.2, 0.5, 0.8, 1.0]) \  
  .addGrid(lr.maxIter, [1, 5, 10]) \  
  .build())
```

Cross Validator

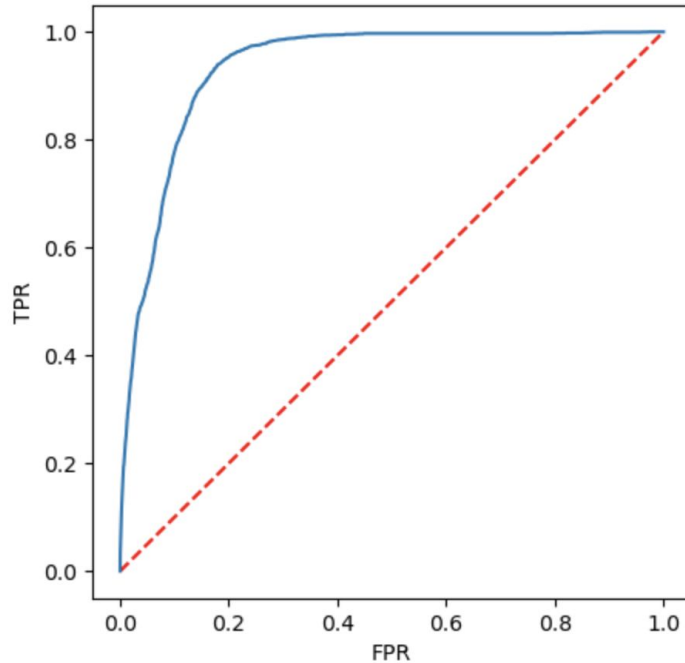
```
= CrossValidator(estimator = lr,  
  estimatorParamMaps = lrParamGrid,  
  evaluator = evaluator,  
  numFolds = 10,  
  seed=314)
```

Best Model

Lambda = 0.01, Max Iterations = 10, Alpha = 0.2

| TP | TN | FP | FN | Accuracy | AUROC |
|-------|-------|-------|-------|----------|---------|
| 0.343 | 0.978 | 0.022 | 0.657 | 0.911114 | 0.66029 |

Model Construction | Logistic Regression



| Threshold | TP | TN | FP | FN | Accuracy | AUROC |
|-----------|-------|-------|-------|-------|----------|---------|
| 0.1 | 0.934 | 0.811 | 0.189 | 0.066 | 0.82411 | 0.8727 |
| 0.2 | 0.737 | 0.904 | 0.096 | 0.263 | 0.88638 | 0.82049 |
| 0.3 | 0.554 | 0.943 | 0.057 | 0.446 | 0.90168 | 0.74812 |
| 0.4 | 0.457 | 0.965 | 0.035 | 0.543 | 0.91187 | 0.71108 |
| 0.5 | 0.343 | 0.978 | 0.022 | 0.657 | 0.91114 | 0.66029 |
| 0.6 | 0.211 | 0.988 | 0.012 | 0.789 | 0.90605 | 0.59943 |
| 0.7 | 0.093 | 0.996 | 0.004 | 0.907 | 0.90095 | 0.54468 |
| 0.8 | 0.031 | 0.999 | 0.001 | 0.969 | 0.89694 | 0.51496 |
| 0.9 | 0.003 | 1.0 | 0.0 | 0.997 | 0.89512 | 0.50173 |

Model Construction | Random Forest

Parameter Grid

```
= (ParamGridBuilder() \  
  .addGrid(rf.maxDepth, [2, 5, 10]) \  
  .addGrid(rf.maxBins, [10, 20]) \  
  .addGrid(rf.numTrees, [5, 20, 50]) \  
  .build())
```

Cross Validator

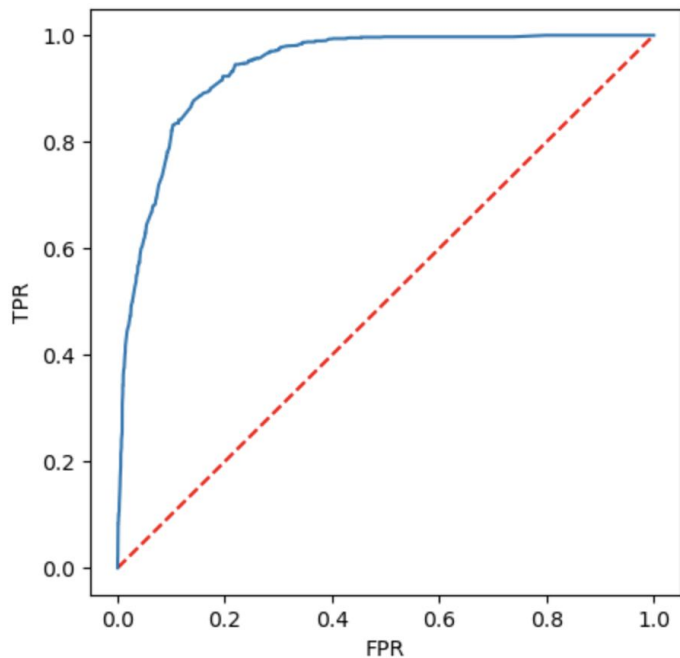
```
= CrossValidator(estimator = rf,  
  estimatorParamMaps = rfParamGrid,  
  evaluator = evaluator,  
  numFolds = 10,  
  seed=314)
```

Best Model

Max Depth = 10, Max Bins = 10, Num Trees = 5

| TP | TN | FP | FN | Accuracy | AUROC |
|-------|--------|--------|-------|----------|---------|
| 0.083 | 0.9996 | 4.0E-4 | 0.917 | 0.90313 | 0.54132 |

Model Construction | Random Forest



| Threshold | TP | TN | FP | FN | Accuracy | AUROC |
|-----------|--------|--------|--------|--------|----------|---------|
| 0.2 | 0.7889 | 0.8596 | 0.1404 | 0.2111 | 0.85215 | 0.82426 |
| 0.1 | 0.9827 | 0.6227 | 0.3773 | 0.0173 | 0.6606 | 0.8027 |
| 0.3 | 0.5571 | 0.9573 | 0.0427 | 0.4429 | 0.91515 | 0.75718 |
| 0.4 | 0.1972 | 0.9923 | 0.0077 | 0.8028 | 0.90859 | 0.59475 |
| 0.5 | 0.083 | 0.9996 | 4.0E-4 | 0.917 | 0.90313 | 0.54132 |
| 0.6 | 0.0415 | 1.0 | 0.0 | 0.9585 | 0.89913 | 0.52076 |
| 0.7 | 0.0173 | 1.0 | 0.0 | 0.9827 | 0.89658 | 0.50865 |
| 0.8 | 0.0 | 1.0 | 0.0 | 1.0 | 0.89476 | 0.5 |
| 0.9 | 0.0 | 1.0 | 0.0 | 1.0 | 0.89476 | 0.5 |

Model Construction | Decision Trees

Parameter Grid

```
= (ParamGridBuilder() \  
  .addGrid(dt.maxDepth, [2, 5, 10]) \  
  .addGrid(dt.maxBins, [10, 20]) \  
  .build())
```

Cross Validator

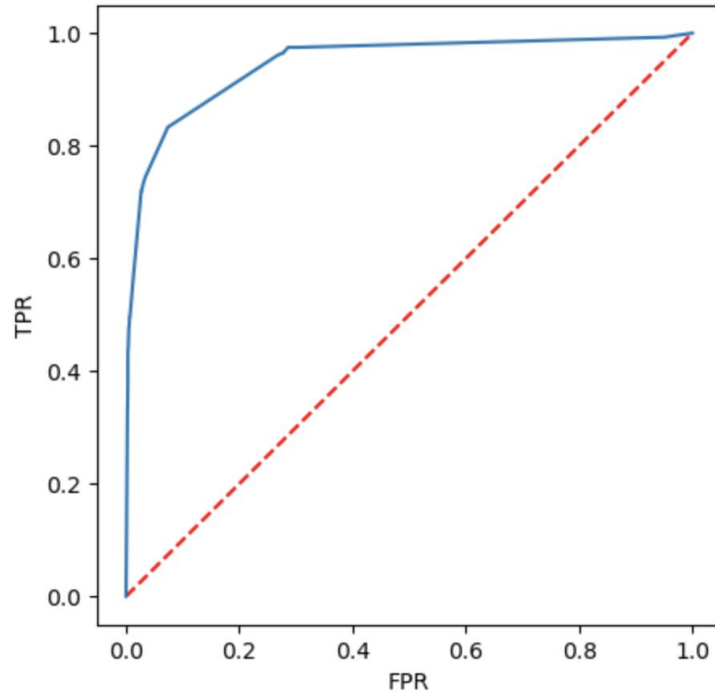
```
= CrossValidator(estimator = dt,  
  estimatorParamMaps = dtParamGrid,  
  evaluator = evaluator,  
  numFolds = 10,  
  seed=314)
```

Best Model

Max Depth = 10, Max Bins = 20

| TP | TN | FP | FN | Accuracy | AUROC |
|--------|-------|-------|--------|----------|--------|
| 0.6886 | 0.978 | 0.022 | 0.3114 | 0.94756 | 0.8333 |

Model Construction | Decision Trees



| Threshold | TP | TN | FP | FN | Accuracy | AUROC |
|-----------|-------|-------|-------|-------|----------|---------|
| 0.1 | 0.803 | 0.901 | 0.099 | 0.197 | 0.89039 | 0.85173 |
| 0.2 | 0.699 | 0.974 | 0.026 | 0.301 | 0.94465 | 0.83625 |
| 0.3 | 0.692 | 0.978 | 0.022 | 0.308 | 0.94756 | 0.83483 |
| 0.4 | 0.689 | 0.978 | 0.022 | 0.311 | 0.94756 | 0.8333 |
| 0.5 | 0.689 | 0.978 | 0.022 | 0.311 | 0.94756 | 0.8333 |
| 0.6 | 0.498 | 0.994 | 0.006 | 0.502 | 0.9421 | 0.74629 |
| 0.7 | 0.484 | 0.996 | 0.004 | 0.516 | 0.9421 | 0.74018 |
| 0.8 | 0.474 | 0.996 | 0.004 | 0.526 | 0.94101 | 0.73499 |
| 0.9 | 0.415 | 0.998 | 0.002 | 0.585 | 0.937 | 0.7068 |

Model Evaluation

| Model | Threshold | TP | TN | FP | FN | Accuracy | AUROC |
|-------|-----------|-------|-------|-------|-------|----------|---------|
| LR | 0.4 | 0.457 | 0.965 | 0.035 | 0.543 | 0.91187 | 0.71108 |
| RF | 0.4 | 0.197 | 0.992 | 0.008 | 0.803 | 0.90859 | 0.59475 |
| DT | 0.5 | 0.689 | 0.978 | 0.022 | 0.311 | 0.94756 | 0.8333 |

Model Evaluation

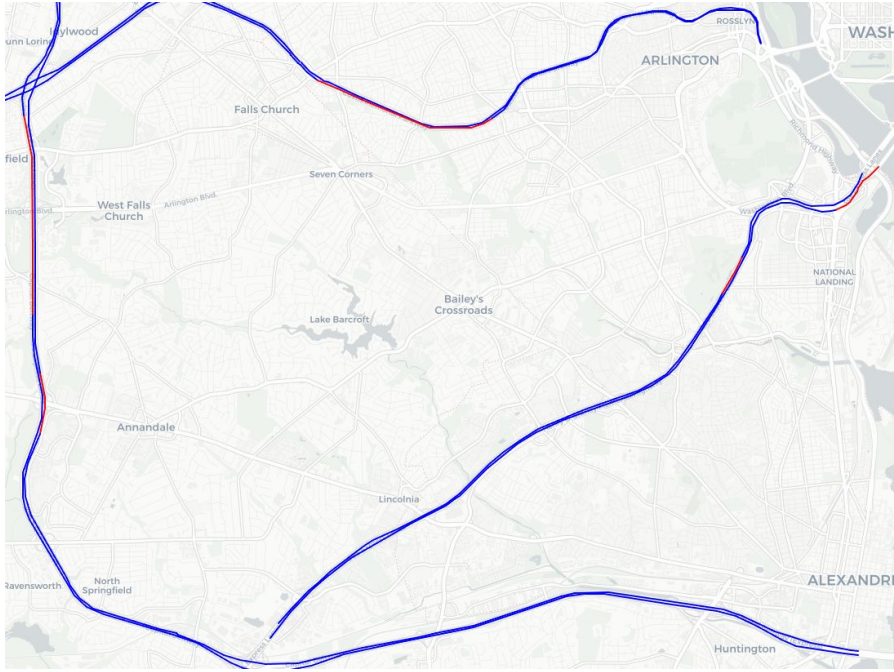
| prediction | count |
|------------|-------|
| 0.0 | 25818 |
| 1.0 | 1494 |

| YEAR | prediction | count |
|------|------------|-------|
| 2021 | 0.0 | 6458 |
| 2021 | 1.0 | 370 |
| 2022 | 0.0 | 6453 |
| 2022 | 1.0 | 375 |
| 2023 | 0.0 | 6453 |
| 2023 | 1.0 | 375 |
| 2024 | 0.0 | 6454 |
| 2024 | 1.0 | 374 |

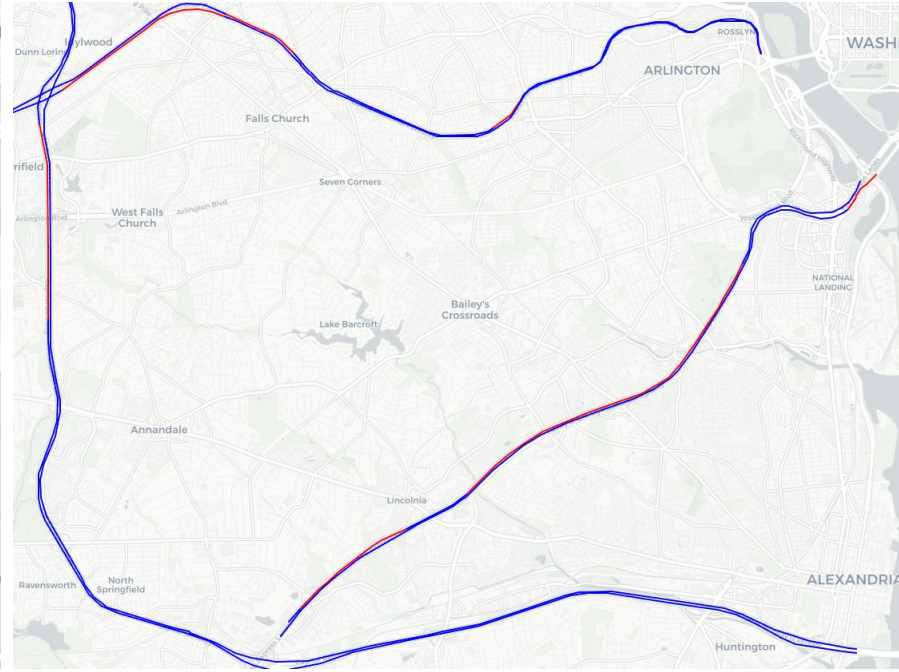
| Year | sum(prediction) | count | |
|------|-----------------|-------|------------------|
| 2021 | 0.0 | 1571 | |
| 2021 | 1.0 | 34 | |
| 2021 | 2.0 | 12 | → 136 unreliable |
| 2021 | 3.0 | 24 | |
| 2021 | 4.0 | 66 | |
| 2022 | 0.0 | 1573 | |
| 2022 | 1.0 | 32 | |
| 2022 | 2.0 | 11 | → 134 unreliable |
| 2022 | 3.0 | 25 | |
| 2022 | 4.0 | 66 | |
| 2023 | 0.0 | 1573 | |
| 2023 | 1.0 | 34 | |
| 2023 | 2.0 | 9 | → 134 unreliable |
| 2023 | 3.0 | 25 | |
| 2023 | 4.0 | 66 | |
| 2024 | 0.0 | 1569 | |
| 2024 | 1.0 | 37 | |
| 2024 | 2.0 | 6 | → 138 unreliable |
| 2024 | 3.0 | 29 | |
| 2024 | 4.0 | 66 | |

Model Evaluation - NOVA

2020 Historical

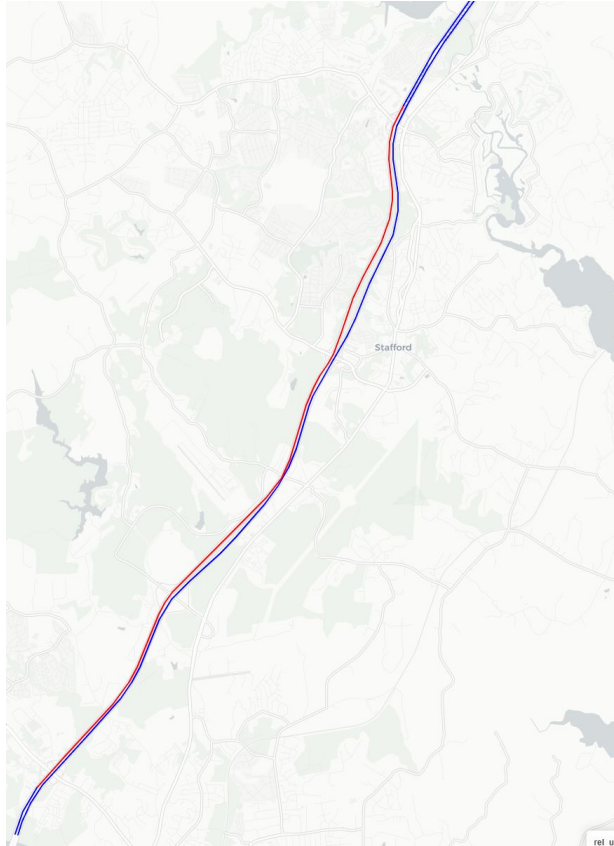


2021 Forecasted

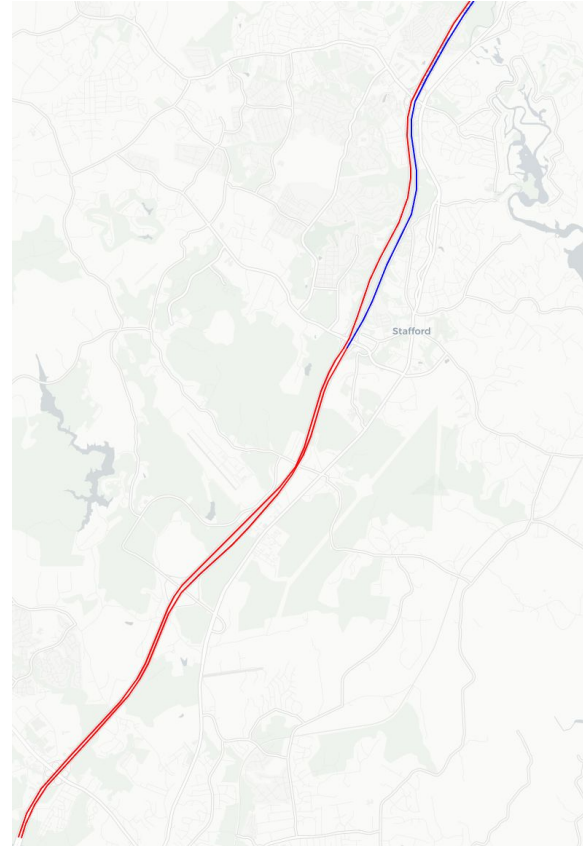


Model Evaluation - Stafford Area

2020 Historical

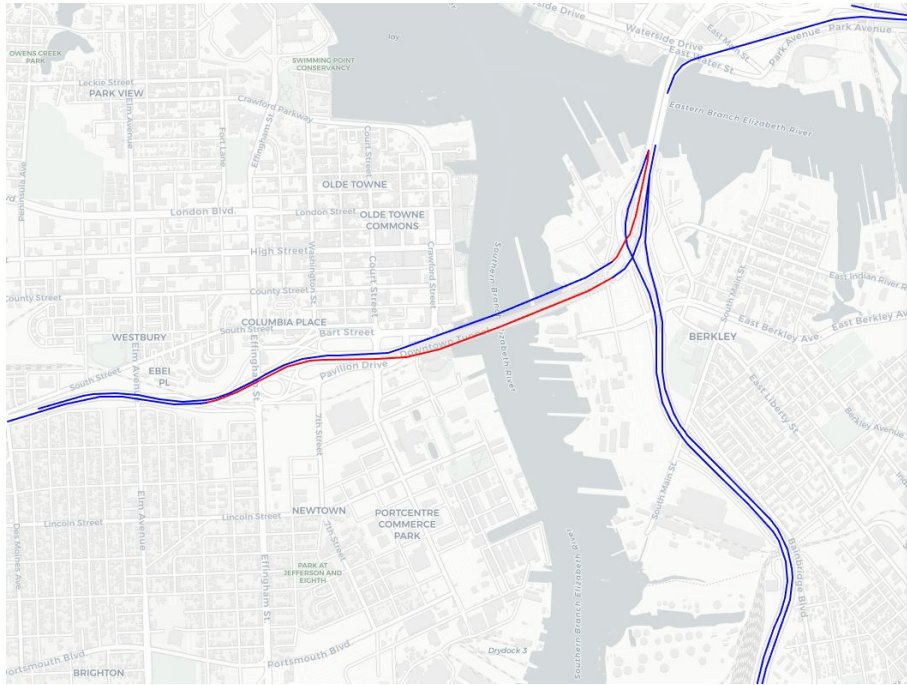


2021 Forecasted

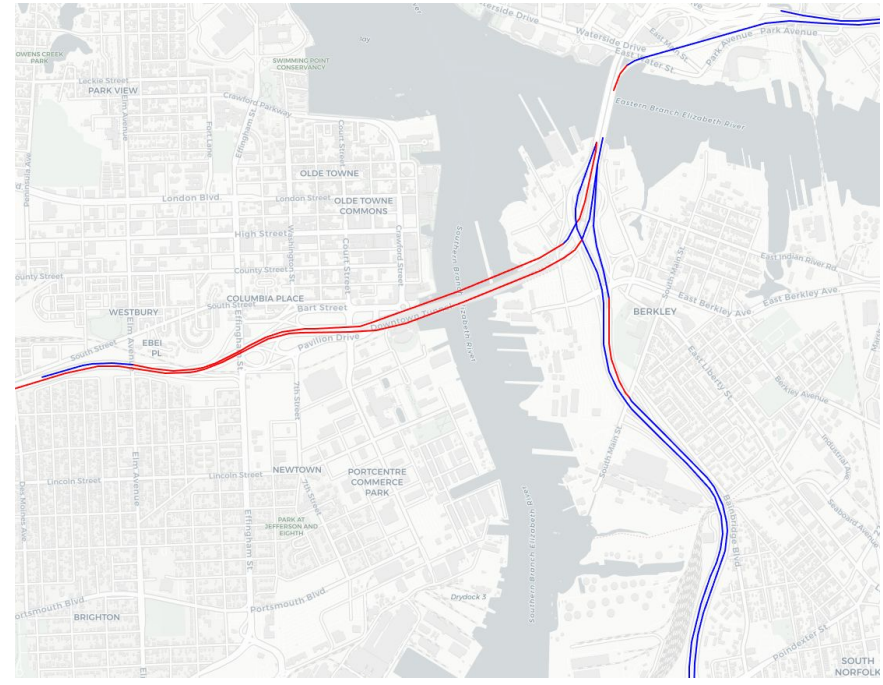


Model Evaluation - Portsmouth

2020 Historical



2021 Forecasted



Conclusions & Future Work

- The Decision Tree model projected on average 136 highway segments to be unreliable each year from 2021-2024
- Tradeoffs based on context of the problem
- Future Work: Evaluating the methodology used for the Forecasted data set
- Future Work: Developing methods to advise the state on which enhancements would make specific unreliable segments more reliable.